



OIPE

RAW SEQUENCE LISTING

DATE: 05/21/2002

PATENT APPLICATION: US/09/995,007

TIME: 14:23:16

Input Set : N:\Crif3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: NANCY CARRASCO, ET AL.

7 (ii) TITLE OF INVENTION: THYROID SODIUM/IODIDE SYMPORTER AND
8 NUCLEIC ACID ENCODING SAME

10 (iii) NUMBER OF SEQUENCES: 2

12 (iv) CORRESPONDENCE ADDRESS:

13 (A) ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

14 (B) STREET: 90 PARK AVENUE

15 (C) CITY: NEW YORK

16 (D) STATE: NEW YORK

17 (E) COUNTRY: U.S.A.

18 (F) ZIP: 10016

20 (v) COMPUTER READABLE FORM:

21 (A) MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
22 DISKETTE

23 (B) COMPUTER: IBM PC COMPATIBLE

24 (C) OPERATING SYSTEM: MS-DOS

25 (D) SOFTWARE: ASCII

27 (vi) CURRENT APPLICATION DATA:

C--> 28 (A) APPLICATION NUMBER: US/09/995,007

C--> 29 (B) FILING DATE: 26-Nov-2001

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: US/08/595,553A

34 (B) FILING DATE: FEBRUARY 1, 1996

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: CRAIG J. ARNOLD

40 (B) REGISTRATION NUMBER: 34,287

41 (C) REFERENCE/DOCKET NUMBER: 96700/393

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: (212) 697-5995

45 (B) TELEFAX: (212) 286-0854 or 286-0082

46 (C) TELEX: TWX 710-581-4766

49 (2) INFORMATION FOR SEQ ID NO: 1:

51 (i) SEQUENCE CHARACTERISTICS:

52 (A) LENGTH: 2839

53 (B) TYPE: NUCLEIC ACID

54 (C) STRANDEDNESS: DOUBLE

55 (D) TOPOLOGY: LINEAR

W--> 57 (ii) MOLECULE TYPE:

58 (A) DESCRIPTION: OLIGONUCLEOTIDE

60 (iii) HYPOTHETICAL: NO

62 (iv) ANTI-SENSE: NO

ENTERED

RAW SEQUENCE LISTING

DATE: 05/21/2002

PATENT APPLICATION: US/09/995,007

TIME: 14:23:16

Input Set : N:\Crif3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

```

64      (vi) ORIGINAL SOURCE:
65      (A) ORGANISM: RAT
C--> 66      (C) INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER
68      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
70 GAATTCCGGG TCGACCACGC GTCCGGCGGT GACTCGCGCT GCGACTCTCC      50
72 CACTGACCGA GAGTCCCCGA CGTCCTCCGC ATCCTCTCCT CACCGAGTCA      100
74 CCTGTCTCC  ATG GAG GGT GCG GAG GCC GGG GCC CGG GCC ACC      142
75      Met Glu Gly Ala Glu Ala Gly Ala Arg Ala Thr
76      1          5          10
78 TTC GGC GCC TGG GAC TAC GGC GTG TTC GCG ACC ATG CTG CTG      184
80 Phe Gly Ala Trp Asp Tyr Gly Val Phe Ala Thr Met Leu Leu
81      15          20          25
83 GTG TCC ACG GGC ATC GGG CTA TGG GTC GGC CTG GCC CGC GGT      226
85 Val Ser Thr Gly Ile Gly Leu Trp Val Gly Leu Ala Arg Gly
86      30          35
88 GGC CAA CGC AGT GCC GAC GAC TTC TTT ACC GGG GGC CGG CAG      268
89 Gly Gln Arg Ser Ala Asp Asp Phe Phe Thr Gly Gly Arg Gln
90 40          45          50
92 TTG GCA GCC GTT CCT GTG GGG CTG TCG CTG GCC GCC AGT TTC      310
93 Leu Ala Ala Val Pro Val Gly Leu Ser Leu Ala Ala Ser Phe
94      55          60          65
96 ATG TCG GCT GTG CAG GTG CTC GGG GTC CCC GCC GAG GCA GCG      352
97 Met Ser Ala Val Gln Val Leu Gly Val Pro Ala Glu Ala Ala
98      70          75          80
100 CGC TAC GGG CTC AAG TTC CTG TGG ATG TGC GCG GGT CAG TTG      394
101 Arg Tyr Gly Leu Lys Phe Leu Trp Met Cys Ala Gly Gln Leu
102      85          90          95
104 CTC AAC TCG CTG CTC ACA GCG TTT CTC TTC TTG CCG ATC TTC      436
105 Leu Asn Ser Leu Leu Thr Ala Phe Leu Phe Leu Pro Ile Phe
106      100          105
108 TAC CGC CTG GGC CTT ACC AGC ACC TAC CAG TAC CTA GAG CTG      478
109 Tyr Arg Leu Gly Leu Thr Ser Thr Tyr Gln Tyr Leu Glu Leu
110 110          115          120
112 CGC TTC AGC CGA GCG GTC CGG CTC TGC GGG ACG CTG CAG TAC      520
113 Arg Phe Ser Arg Ala Val Arg Leu Cys Gly Thr Leu Gln Tyr
114      125          130          135
116 TTG GTG GCC ACG ATG CTG TAT ACA GGC ATC GTG ATC TAC GCG      562
117 Leu Val Ala Thr Met Leu Tyr Thr Gly Ile Val Ile Tyr Ala
118      140          145          150
120 CCT GCG CTC ATC CTG AAC CAA GTG ACC GGG TTG GAC ATC TGG      604
121 Pro Ala Leu Ile Leu Asn Gln Val Thr Gly Leu Asp Ile Trp
122      155          160          165
124 GCA TCG CTC CTG TCC ACA GGA ATC ATC TGC ACC TTG TAC ACT      646
125 Ala Ser Leu Leu Ser Thr Gly Ile Ile Cys Thr Leu Tyr Thr
126      170          175
128 ACC GTG GGT GGT ATG AAG GCC GTG GTC TGG ACA GAT GTG TTC      688
129 Thr Val Gly Gly Met Lys Ala Val Val Trp Thr Asp Val Phe
130 180          185          190
132 CAG GTT GTG GTA ATG CTC GTT GGC TTC TGG GTG ATC CTG GCC      730

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/995,007

DATE: 05/21/2002

TIME: 14:23:16

Input Set : N:\Crif3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

133	Gln	Val	Val	Val	Met	Leu	Val	Gly	Phe	Trp	Val	Ile	Leu	Ala	
134		195					200					205			
136	CGA	GGC	GTC	ATT	CTC	CTG	GGG	GGT	CCC	CGG	AAC	GTG	CTC	AGC	772
137	Arg	Gly	Val	Ile	Leu	Leu	Gly	Gly	Pro	Arg	Asn	Val	Leu	Ser	
138			210				215					220			
140	CTC	GCT	CAG	AAC	CAT	TCC	CGG	ATC	AAC	CTG	ATG	GAC	TTT	GAC	814
141	Leu	Ala	Gln	Asn	His	Ser	Arg	Ile	Asn	Leu	Met	Asp	Phe	Asp	
142			225					230				235			
144	CCT	GAT	CCT	CGG	AGC	CGG	TAC	ACC	TTC	TGG	ACT	TTC	ATA	GTG	856
145	Pro	Asp	Pro	Arg	Ser	Arg	Tyr	Thr	Phe	Trp	Thr	Phe	Ile	Val	
146			240					245							
148	GGT	GGC	ACA	CTG	GTG	TGG	CTC	TCC	ATG	TAC	GGT	GTG	AAC	CAA	898
149	Gly	Gly	Thr	Leu	Val	Trp	Leu	Ser	Met	Tyr	Gly	Val	Asn	Gln	
150	250					255					260				
152	GCC	CAG	GTA	CAG	CGC	TAT	GTG	GCC	TGC	CAC	ACA	GAG	GGA	AAG	940
153	Ala	Gln	Val	Gln	Arg	Tyr	Val	Ala	Cys	His	Thr	Glu	Gly	Lys	
154		265				270					275				
156	GCC	AAA	CTG	GCC	CTG	CTT	GTC	AAC	CAG	CTG	GGC	CTC	TTC	CTG	982
157	Ala	Lys	Leu	Ala	Leu	Leu	Val	Asn	Gln	Leu	Gly	Leu	Phe	Leu	
158		280				285					290				
160	ATT	GTG	GCC	AGT	GCA	GCT	TGC	TGT	GGC	ATT	GTC	ATG	TTC	GTC	1024
161	Ile	Val	Ala	Ser	Ala	Ala	Cys	Cys	Gly	Ile	Val	Met	Phe	Val	
162			295					300				305			
164	TAC	TAC	AAG	GAC	TGT	GAC	CCC	CTC	CTC	ACA	GGC	CGT	ATC	TCA	1066
165	Tyr	Tyr	Lys	Asp	Cys	Asp	Pro	Leu	Leu	Thr	Gly	Arg	Ile	Ser	
166			310					315							
168	GCC	CCC	GAC	CAG	TAC	ATG	CCG	CTG	CTT	GTG	TTG	GAC	ATT	TTT	1108
169	Ala	Pro	Asp	Gln	Tyr	Met	Pro	Leu	Leu	Val	Leu	Asp	Ile	Phe	
170	320				325					330					
172	GAG	GAT	CTG	CCC	GGG	GTC	CCC	GGG	CTC	TTC	CTG	GCC	TGT	GCC	1150
173	Glu	Asp	Leu	Pro	Gly	Val	Pro	Gly	Leu	Phe	Leu	Ala	Cys	Ala	
174		335				340					345				
176	TAC	AGT	GGC	ACC	CTC	AGC	ACT	GCA	TCC	ACC	AGC	ATC	AAC	GCC	1192
177	Tyr	Ser	Gly	Thr	Leu	Ser	Thr	Ala	Ser	Thr	Ser	Ile	Asn	Ala	
178		350				355					360				
180	ATG	GCA	GCT	GTG	ACT	GTG	GAA	GAC	CTC	ATC	AAG	CCG	AGG	ATG	1234
181	Met	Ala	Ala	Val	Thr	Val	Glu	Asp	Leu	Ile	Lys	Pro	Arg	Met	
182			365				370				375				
184	CCT	GGC	CTG	GCA	CCT	CGG	AAG	TTG	GTT	TTC	ATC	TCT	AAA	GGG	1276
185	Pro	Gly	Leu	Ala	Pro	Arg	Lys	Leu	Val	Phe	Ile	Ser	Lys	Gly	
186			380					385							
188	CTC	TCA	TTC	ATC	TAC	GGC	TCT	GCC	TGC	CTC	ACT	GTG	GCT	GCT	1318
189	Leu	Ser	Phe	Ile	Tyr	Gly	Ser	Ala	Cys	Leu	Thr	Val	Ala	Ala	
190	390				395					400					
192	CTG	TCC	TCA	CTG	CTG	GGA	GGT	GGT	GTC	CTC	CAG	GGT	TCC	TTC	1360
193	Leu	Ser	Ser	Leu	Leu	Gly	Gly	Gly	Val	Leu	Gln	Gly	Ser	Phe	
194		405				410					415				
196	ACT	GTG	ATG	GGT	GTC	ATC	AGT	GGG	CCT	CTA	CTA	GGC	GCC	TTC	1402
197	Thr	Val	Met	Gly	Val	Ile	Ser	Gly	Pro	Leu	Leu	Gly	Ala	Phe	

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/995,007

DATE: 05/21/2002

TIME: 14:23:16

Input Set : N:\Crif3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

198	420	425	430	
200	ACG CTT GGG ATG CTG CTC CCA GCC TGC AAC ACG CCA GGC GTT	1444		
201	Thr Leu Gly Met Leu Leu Pro Ala Cys Asn Thr Pro Gly Val			
202	435 440 445			
204	CTC TCC GGG TTG GCA GCA GGC TTG GCT GTA TCC CTG TGG GTG	1486		
205	Leu Ser Gly Leu Ala Ala Gly Leu Ala Val Ser Leu Trp Val			
206	450 455			
208	GCC GTA GGG GCC ACA CTG TAT CCC CCT GGA GAG CAG ACC ATG	1528		
209	Ala Val Gly Ala Thr Leu Tyr Pro Pro Gly Glu Gln Thr Met			
210	460 465 470			
212	GGG GTG CTG CCC ACC TCG GCT GCA GGC TGC ACC AAC GAT TCG	1570		
213	Gly Val Leu Pro Thr Ser Ala Ala Gly Cys Thr Asn Asp Ser			
214	475 480 485			
216	GTC CTC CTG GGC CCA CCT GGA GCC ACC AAC GCT TCC AAC GGG	1612		
217	Val Leu Leu Gly Pro Pro Gly Ala Thr Asn Ala Ser Asn Gly			
218	490 495 500			
220	ATC CCC AGT TCT GGA ATG GAC ACG GGC CGC CCT GCC CTC GCT	1654		
221	Ile Pro Ser Ser Gly Met Asp Thr Gly Arg Pro Ala Leu Ala			
222	505 510 515			
224	GAT ACC TTT TAC GCC ATC TCC TAT CTC TAT TAC GGG GCT CTG	1696		
225	Asp Thr Phe Tyr Ala Ile Ser Tyr Leu Tyr Tyr Gly Ala Leu			
226	520 525			
228	GGC ACG CTG ACC ACC ATG CTT TGC GGT GCT CTC ATC AGC TAC	1738		
229	Gly Thr Leu Thr Thr Met Leu Cys Gly Ala Leu Ile Ser Tyr			
230	530 535 540			
232	CTT ACT GGT CCC ACC AAG CGC AGC TCC CTG GGT CCC GGA TTG	1780		
233	Leu Thr Gly Pro Thr Lys Arg Ser Ser Leu Gly Pro Gly Leu			
234	545 550 555			
236	CTG TGG TGG GAC CTT GCT CGA CAG ACA GCG TCT GTG GCC CCA	1822		
237	Leu Trp Trp Asp Leu Ala Arg Gln Thr Ala Ser Val Ala Pro			
238	560 565 570			
240	AAG GAA GAC ACT GCC ACC CTG GAG GAG AGC CTG GTG AAG GGA	1864		
241	Lys Glu Asp Thr Ala Thr Leu Glu Glu Ser Leu Val Lys Gly			
242	575 580 585			
244	CCG GAA GAC ATC CCT GCT GTG ACC AAG AAG CCC CCT GGC CTC	1906		
245	Pro Glu Asp Ile Pro Ala Val Thr Lys Lys Pro Pro Gly Leu			
246	590 595			
248	AAG CCA GGC GCC GAG ACC CAC CCC CTG TAT CTG GGG CAC GAT	1948		
249	Lys Pro Gly Ala Glu Thr His Pro Leu Tyr Leu Gly His Asp			
250	600 605 610			
252	GTG GAG ACC AAC CTC TGAGGGCGGG GTCCAAGAAG GCCAATCACA	1993		
253	Val Glu Thr Asn Leu			
254	615			
256	GGCCTCGGGC CAGCAGCCTC CTCTCTGGAT GGTTGGACCT GAGCATATAT	2043		
258	AGAAGCTTGG CTGATACATG CCCTGCCCAG AAGTCCCTGT GTCTTACCCG	2093		
260	CACCAAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG	2143		
262	GAGTTGGTTC TCCATCCACA AAGGAAACCG TCTGGAACCT TCATGCCCTT	2193		
264	GTAGATTTC GTAGGCAGCG GAGAACAATC AGCTTCTCCA GACTGAGGTT	2243		
266	TTCTCATTTA TCAGGCAGAG AAACGGAGGG CTGTCACCCC AACACCGGGG	2293		

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/995,007

DATE: 05/21/2002

TIME: 14:23:16

Input Set : N:\Crif3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

```

268 AGGAGACAGT AGAAGGGTCA TAGATACAAA GAAAACTAAG GCAGAGGGAG      2343
270 AAATGAATTG TCTACAGAGC ACAGAGCTCC AAGGATTGTG AAGCTACCTT      2393
272 GAGGTGCCAA GGGACGGATT CTCAGAGCCT TCACAAGACA CAAACGGACG      2443
274 AGTTGCCTCC TCCAATTCAG ATGGTTTGCA GACTATCAGA GAACATGTTT      2493
276 CTCCTGTGAT CAGCTACCTA GCCTCTGCCA ACGTGTTCCT GCTTCCAGGA      2543
278 GGCCACACAG ACCCCACCCC CCATGCTCTC ACCCTTTACC CCTGTGCTTT      2593
280 TCACACACTA GGCAACTGCT CCACCACAGG ACCTCACACC TAGACCTCCG      2643
282 TTTTGTACAC AGGGCCTTAA GGTAATCTGG CTGCCATCTG ACTATCTCTC      2693
284 AGCACGTTCA CGTGTACAAT ATTTTCATTCT TTTTCATTGC CAAGTTGTCT      2743
286 TGTAAGGAGA GACCACAATG TGTCATCCAT GCCCAGCTTT TGTGTCTAAC      2793
288 AAATAAAATC GCTGAAGGTG TTCAGGTGCA ATGGCCTGTG ACATTA      2839
291 (2) INFORMATION FOR SEQ ID NO: 2:
293     (i) SEQUENCE CHARACTERISTICS:
294         (A) LENGTH: 618
295         (B) TYPE: AMINO ACID
296         (D) TOPOLOGY: LINEAR
W--> 298     (ii) MOLECULE TYPE:
299         (A) DESCRIPTION: PROTEIN
301     (iii) HYPOTHETICAL: YES
303     (vi) ORIGINAL SOURCE:
304         (A) ORGANISM: RAT
C--> 305     (C) INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER
307     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
311         Met Glu Gly Ala Glu Ala Gly Ala Arg Ala Thr
312             1             5             10
314 Phe Gly Ala Trp Asp Tyr Gly Val Phe Ala Thr Met Leu Leu
315             15             20             25
317 Val Ser Thr Gly Ile Gly Leu Trp Val Gly Leu Ala Arg Gly
318             30             35
320 Gly Gln Arg Ser Ala Asp Asp Phe Phe Thr Gly Gly Arg Gln
321 40             45             50
323 Leu Ala Ala Val Pro Val Gly Leu Ser Leu Ala Ala Ser Phe
324 55             60             65
326 Met Ser Ala Val Gln Val Leu Gly Val Pro Ala Glu Ala Ala
327 70             75             80
329 Arg Tyr Gly Leu Lys Phe Leu Trp Met Cys Ala Gly Gln Leu
330 85             90             95
332 Leu Asn Ser Leu Leu Thr Ala Phe Leu Phe Leu Pro Ile Phe
333 100            105
335 Tyr Arg Leu Gly Leu Thr Ser Thr Tyr Gln Tyr Leu Glu Leu
336 110            115            120
338 Arg Phe Ser Arg Ala Val Arg Leu Cys Gly Thr Leu Gln Tyr
339 125            130            135
341 Leu Val Ala Thr Met Leu Tyr Thr Gly Ile Val Ile Tyr Ala
342 140            145            150
344 Pro Ala Leu Ile Leu Asn Gln Val Thr Gly Leu Asp Ile Trp
345 155            160            165
347 Ala Ser Leu Leu Ser Thr Gly Ile Ile Cys Thr Leu Tyr Thr
348 170            175

```

VERIFICATION SUMMARY

DATE: 05/21/2002

PATENT APPLICATION: US/09/995,007

TIME: 14:23:17

Input Set : N:\Crf3\RULE60\09995007.raw

Output Set: N:\CRF3\05212002\I995007.raw

L:28 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:29 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:66 M:220 C: Keyword misspelled or invalid format, [(C) INDIVIDUAL ISOLATE:]
L:57 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=1
L:305 M:220 C: Keyword misspelled or invalid format, [(C) INDIVIDUAL ISOLATE:]
L:298 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=2